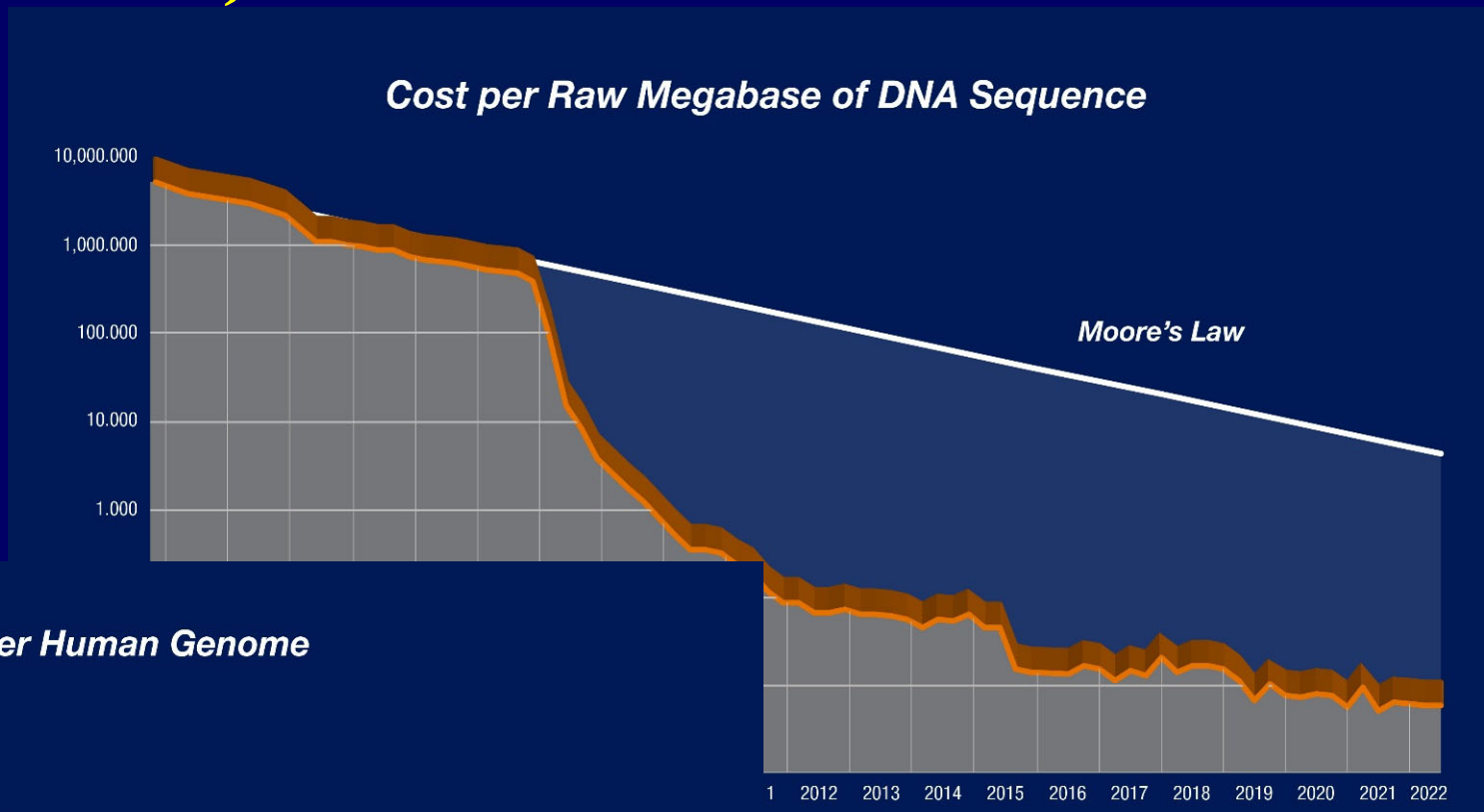
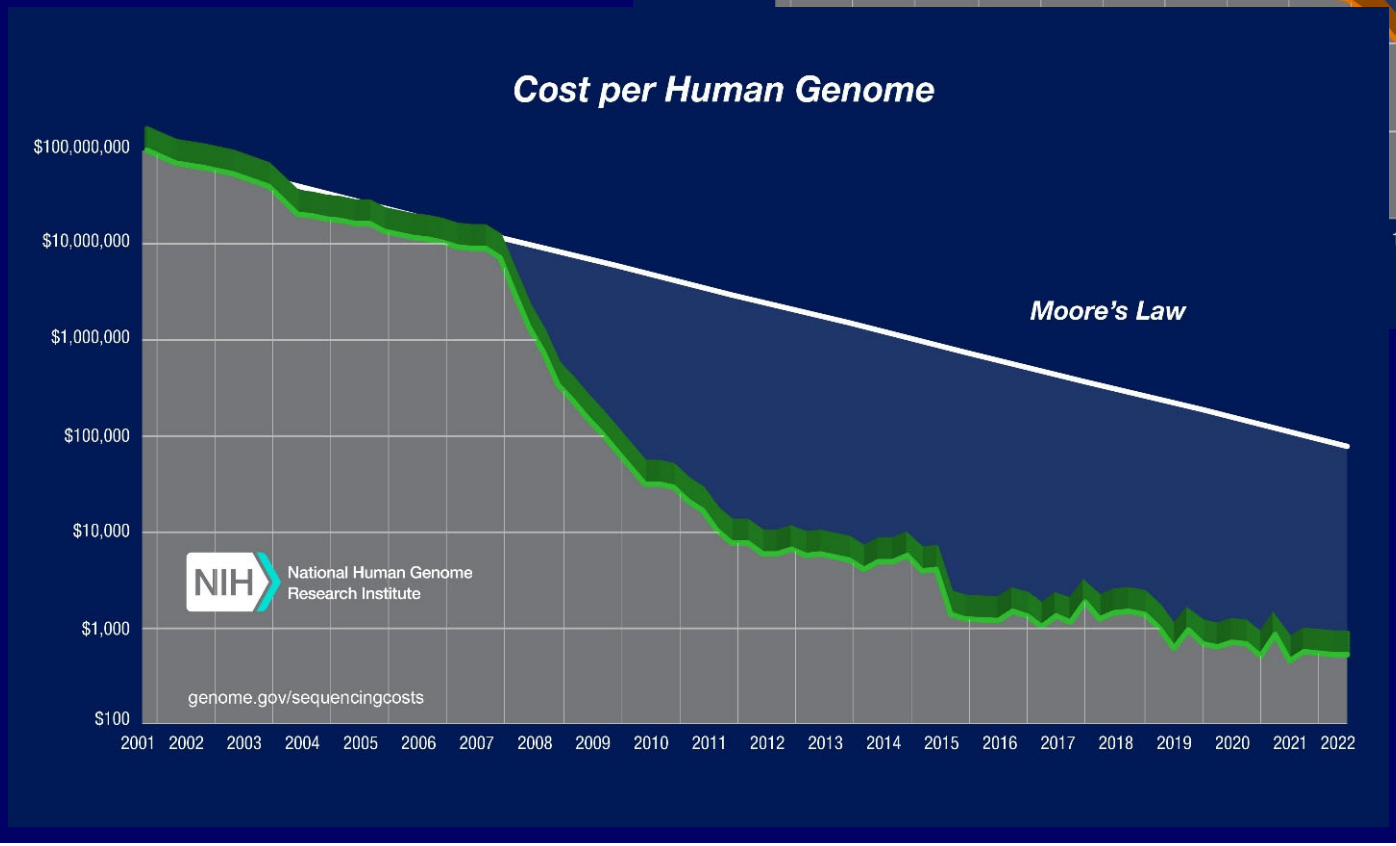
HTS (WGS) data





Date	Cost per Mb	Cost per Genome
Feb-22	$0.006	$525
May-22	$0.006	$525

HTS (WGS) are used for:

- Expression profiling: RNA-Seq, scRNA-Seq, etc.
- DNA accessibility: ATAC-Seq, scATAC-Seq, DNase-Seq, FAIRE-Seq, etc.
- Histone Modification: ChIP-Seq, scChIP-Seq, etc.
- Nuclear organization: Hi-C, 4C, 3C, ChIA-PET, etc.
- Detecting mutations/variations, ….

General processing of HTS data

1. Mapping to the targeted genome.
2. Quantification.
3. Comparison.
4. Further analysis (GO enrichment assay, etc.)

HTS Analysis (I).

Hardware considerations for typical small RNA-Seq / ChIP-Seq experiments (4-16 samples):

- Storage: 1-2 TB
- RAM: >32GB (depending on the size of FASTQ file that your will be handling).
- CPU core: at least 4, preferably 16 or more.
- Operating system: Linux
- Software: GPL (open software) applications available for most analysis.

HTS Analysis (II).

General setup consideration

- A designated project folder in HiPerGator
  - Large data files, computation-intensive processing.
  - For this class, consider performing RNA-Seq in your course folder.
- A corresponding folder in local computer
  - Scripts, results (spreadsheets), figures, etc.
- Run analysis

HTS – getting organized in HiPerGator

Practice:

1. Log in to HiPerGator with OOD.
2. Open a terminal in your GMS6014/ folder.
3. Use “ls -l” to observe user rights of your folder and other folders.
4. Use “chmod –R 744 foldername” to modify accessibility
5. Try to access again

Mapping HTS (WGS) reads to genome

- Choose your genome.
  - Advantage of using the same genome release within a group.
- Different strategies of mapping.
  - Splice-aware vs. -unaware

Mapping HTS (WGS) reads to genome

Observe: STAR mapping algorithm.

Q: What type of aligner would you use for ChIP-Seq?

- Splice-aware
- Splice-unaware

Mapping HTS (WGS) reads to genome

Observe: GTF (Gene Transfer Format) file .

Practice: mapping RNA-Seq reads with STAR aligner.

Generate index from .fasta and .gtf

```
#!/bin/sh
#SBATCH --job-name=Dm6_Star_Index
#SBATCH --mail-type=ALL
#SBATCH --mail-user=leizhou@ufl.edu
#SBATCH --mem-per-cpu=6gb
#SBATCH --cpus-per-task=8
#SBATCH --qos=zhou
#SBATCH -t 3:00:00
#SBATCH --output=STAR_Index_%j.log

module load star

STAR --runThreadN 16 \
--runMode genomeGenerate \
--genomeDir Dm6.67.StarIndex \
--genomeFastaFiles ./Dm6.67.fa \
--sjdbGTFfile ./Dm6.67.gtf \
--sjdbOverhang 99
```